Guidelines for Development of a DATA MANAGEMENT PLAN (DMP)

Earth Science Division
NASA Science Mission Directorate

National Aeronautics and Space Administration

Version 1.0
January 2011

Contents

Preface

The purpose of the Data Management Plan (DMP) is to address the management of data from Earth science missions, from the time of their data collection/observation, to their entry into permanent archives. Unless otherwise specified, the term "project" is used below to refer to the entity responsible for the missions or investigations.

Data will typically flow from the project's sensors through "institutional facilities" not "owned" by the project (e.g., telemetry acquisition systems, generic low level data processing systems) to "project-owned" systems and project-funded science teams, and finally from project systems/science teams into archives that are a part of NASA's core data system capabilities. During this data flow, data will be transformed at multiple steps, and not all data may be appropriate for the archives. However, the data are to be treated as national assets and plans must be made for preservation of the appropriate data and making deliberate decisions when data are to be discarded. While NASA is not an agency responsible for permanently archiving Earth science data, it is responsible for ensuring preservation of the data during their active use by NASA-funded science investigations, which may continue well beyond the period of active data collection. Also, NASA is responsible to ensure that the data and associated artifacts are transferred to other agencies that are responsible for permanently archiving the data from NASA missions/investigations. Projects should consider these responsibilities while preparing their data management plans.

The DMP will be used primarily to identify data products, and the NASA and/or non-NASA facilities used during the project life cycle for receiving, processing, archiving and distributing the data and continued management of the data in archives beyond the life of the project. Data in the DMP will be used to support the readiness of these facilities to discharge their responsibilities relative to the project's data. The effectiveness of the data management and archiving process depends on a well-defined relationship between the project, investigators, and the facilities needed to provide support and stewardship for data during and after the life of the project. The creation of the DMP is the first step in establishing this relationship. It is expected that the DMP will be a reference document that will be updated as significant changes occur during the life of the project.

The NASA Procedural Requirements (NPR) 7120.8 [NASA Research and Technology Program and Project Management Requirements (w/change 1 dated 11/24/10)] document states:

- "Data management planning is provided as either a section of the Program Plan, Cross-Program Research Plan, or as a separate document (if applicable). It should address the data or knowledge being developed or captured by the R&T Portfolio Project and define plans for data rights and services. The plan should demonstrate close interaction with the application community, and provide an exit strategy following technology or knowledge transfer. In addition, it should explain how the project will ensure identification, control, and disposition of project records in accordance with NPD 1440.6, NASA Records Management, and NPR 1441.1, Records Retention Schedules. The plan should document how the results of R&T efforts will be disseminated, and it should also document if there are restrictions that limit or prevent the ability to

disseminate data in accordance with NPR 2200.2, Requirements for Documentation, Approval, and Dissemination of NASA Scientific and Technical Information."

Consistent with the above requirement, for each of the Earth science missions, NASA Earth Science Division (ESD) requires the responsible projects to develop a DMP as stated in the Level 1 Requirement below:

- "All terms and conditions of the transfer of data products and associated information to the NASA Earth Science Division (ESD)-assigned data center shall be documented in a Data Management Plan that has been approved by the Earth Science Data and Information System Project."

The ESD-assigned data center referenced above is a part of NASA's core data system capabilities. It is expected that each project develops the DMP in collaboration with the ESD-assigned data centers, but the primary responsibility rests with the project. For simplicity, in the following, the ESD-assigned data center will be referred to as the Science Data Center.

The purpose of this document is to provide guidelines for the DMP to make the assembly of DMPs easier for the projects, and to ensure that DMPs will be in a consistent and useful format. The outline of this document is the recommended outline for all DMPs. In each section, the text describes the type of material which should be included. The format provided here is intended to be general enough to be applied to all Earth science missions. However, modifications can be made to meet the specific needs of a given mission. It is expected that additional documents such as Interface Control Documents (ICDs) and Operations Agreements (OAs) will be developed between the projects and the Science Data Centers and referenced by the DMP. It is also expected that such more detailed documents may require more frequent updates than the DMP.

A DMP is expected to be delivered to NASA SMD's Earth Science Data System Program Executive after approval by the Earth Science Data and Information System (ESDIS) Project at the Goddard Space Flight Center prior to the Investigation Readiness Review (IRR) and at least one update to the document a month before the end of the funded investigation period.

1.0 Introduction

The introduction should include the name of the project, status (phase in project lifecycle as of the date of the DMP or revision thereof) and the purpose of the DMP.

1.2 DMP Development, Maintenance, and Management Responsibility

This section should identify the Project Manager for the mission as being responsible for the DMP. Identification should include location and contact information. The DMP should be delivered to the and to the ESDIS Project two weeks prior to the Investigation Readiness Review (IRR). Following review by SMD, if revisions are required, a revised version should be delivered within a month after the IRR. The DMP will be submitted to a configuration control board (CCB) and kept as official Project documentation for the duration of the Project and transition to the Earth Science Data and Information System (ESDIS) Project's CCB at the end of the Project. The emphasis in the DMP should be on advanced planning of the management of data acquired by NASA throughout the data life cycle.

1.3 Change Control

This section should illustrate the plans for modifications and updates to this document over time, and how those changes will be controlled. It is not expected that this document will change frequently. However, the DMP should be updated when significant changes occur such as changes to operations plans, hardware capabilities or the science data product suite. If there are planned updates, estimated release dates should be given. The change control strategy for the more detailed documents should also be addressed here.

1.4 Relevant Documents

All currently available documents with information relevant to data management for the project should be referenced here, including name and location of source for documents, if not readily available. Such documents should at least include ICD's, Project Plan, mission operations concepts and plans, and relevant documents for each instrument. If some documents are not yet available but planned, this should be indicated, with expected availability dates and organization that will be producing document. If some of the information called for in these guidelines is available in other documents, those documents may be referenced and the material covered very briefly here for completeness and continuity (the intent is to avoid duplication of content and maintenance of redundant documentation).

2.0 Project Overview

An overview of the project should be provided in this section. It should include a summary of the history of the project to its current status, including predecessor or related missions. It should describe the scientific significance of the data and whether the data are intended to demonstrate a capability or to build a collection of measurements over time.

2.1 Project Objectives

The overall objectives of the project should be briefly described here. This should include how the goals of the project, and how the expected results of the project may contribute to some larger goals or objectives.

2.2 Science Objectives

The specific science objectives of the project should be briefly described here. This would include expected results to be gained in certain scientific areas, and could be related to specific instruments. Also addressed here should be who the primary science users are expected to be, and which science objectives are expected to be met initially in the project, and those which may be met through continued accessibility by secondary investigators.

2.3 Mission Summary

A mission summary should be provided in this section, with references to documents where more details can be found. Examples of summary information are:

- type of aircraft, number and approximate dates of flights planned, types of instruments to be flown

2.4 Instrument Overview

An overview description of each of the instruments to be flown should be provided here with references to documents with more detailed information. A separate subsection for each of the instruments is recommended. Examples of summary information are: experimental objectives, type of instrument, data acquisition modes, field of view, resolution, approximate volume of data per day, etc.

3.0 Project Data Flow

An overview of the Project Data Flow should be given here including an overall functional Data Flow Diagram. This diagram should identify the facilities performing various functions as the project progresses through its various mission phases. The diagram should distinguish between project-specific facilities and those that are external to the project (e.g., Science Data Centers, ground stations, and networks). In addition, an overall project timeline should be developed and included, summarizing key milestones and events relevant to data management over the life cycle of the project.

Though each mission is unique, there are fundamental elements, functions, and services that are common. In general, science investigations can be divided into three segments of operation: mission operations, science operations, and post-mission access. During each segment of the project, fundamental functions and services are performed as part of the integrated investigation. Those functions which are project specific should be distinguished from those that are not.

In additional to a top-level functional flow, a timeline of activities and milestones related to data management over the life cycle of the project should be provided. Examples of activities to be included in the timeline are: determination of list of products, determination of data and metadata formats, development of ICDs and OAs with Science Data Centers, interface testing, end-to-end data flow testing, planned start and end of data collection (multiple intervals in case of aircraft missions), start of flow of data to Science Data Centers, start of public access to data (separate dates, if appropriate, for access through project-owned facilities and access through data centers), planned start and end of reprocessing, assembly of archival information package (all items needed to be archived along with data, as the project ends, in order to ensure future understandability of the mission and its data products), etc.

3.1 Mission Operations

A summary of the projects Mission Operations concept should be given in this section. This summary should address data flows for communications and initial processing of telemetry, as well as the mission control and mission planning and scheduling. The facilities associated with each function of these processes should be identified in a diagram, and the distinction made as to which are project specific and which are not.

In addition to the functional flow, a timeline of the mission operations phase should be included. Specifically, the timeline should address the time from mission planning and scheduling to implementation. Issues such as real time or playback data, quick-look data processing, and real-time instrument control should be addressed as appropriate.

It is expected that this section will be a high level summary, and the details will be covered in a separate document referenced here.

3.2 Science Operations

This section should consist of a brief discussion of the project science operations style. The relationship between the principal investigator, individual scientists and their payloads should be described also.

A diagram of the science operations functional flow should be included in this section. In addition, a timeline of the science operations phase should be presented. This should address the responsiveness of the system to the instrument control needs of the scientist.

3.2.1 Science Control

This section should address the following aspects of science control for the project:

- Overall payload Command & Control approach
- Overall payload Health & Safety Monitoring
- Location of facilities
- Specific networks to be used, including any cyber security requirements

The resources required to support decisions regarding the real-time or near-real-time operation of payloads or instruments should be described as well.

3.2.2 Science Planning and Scheduling

The overall approach to science planning and scheduling for the project should be described in this section. If the approach is a distributed system by instrument, then each instrument should be addressed. The write-up should include the location and description of facilities involved.

3.2.3 Science Data Set Generation

This section should address the computing and analysis resources required to produce mission-specific data products. The discussion should address facilities which provide support across all instruments, as well as instrument specific support. Expected latency between observation and product availability to users should be covered in this section.

This section should address how algorithms are developed and validated, how data products are produced, how and why algorithms are updated on an ongoing basis, reprocessing strategy, and what data are required from other sources, not under the control of the project. Also, plans for the preparation and review of Algorithm Theoretical Basis Documents (ATBDs) as well as plans for maintaining them to reflect product versions should be addressed.

In addition, the analysis software that may be used to generate the data sets, and any analysis support used should be identified. This section should also describe project requirements and plans for assuring and documenting data quality including validation and release of products to the archive system.

3.2.4 Project Data Storage and Distribution

The approach to storage and distribution of products may vary from mission to mission, depending on the requirements agreed to between the project and ESD. In some cases, the projects may be required to send data products to Science Data Centers at the outset. In other cases, during the life of the project, the projects may be responsible for the archiving and distribution of their data products. Such products may use project-specific data repositories for this purpose or may have collaborative arrangements with the Science Data Center(s). This section should address the specific approach to be used by the project for storage and distribution of its data products during the life of the mission. It should include the approach and schedule for selection of data and metadata standards that will be used to ensure usability and accessibility of the data products. It should include the timeline for making the standard science data products and associated metadata publicly available including provision of Directory Interchange Format files to the Global Change Master Directory. It should also include an estimate, by product, of data volumes to be stored (e.g., in TB/year). It should describe the strategy for retention of multiple versions of science data products that result from reprocessing. The detailed and more specific information on products will be included and maintained up-to-date in ICDs between the project and the Science Data Center(s).

3.3 Post-Mission Stewardship and Access

All projects are required to ensure that their data products, along with the scientific algorithm software, coefficients, and ancillary data used to generate these products are transferred to the Science Data Center before the end of the project. This will ensure post-mission access to the products and provide capability to retain corporate knowledge and regenerate products if needed in the future. This section should address the activities and functions that are required to ensure the post-mission availability of data and supporting information on a timely basis for use by the science community. It should also indicate anticipated usage by the community after the end of the mission. All standard data products and other associated items to be permanently archived should be identified in this section.

3.3.1 Transition to Science Data Center(s)

This section should identify the Science Data Center(s) to which the data are to be transitioned and address how data will transition from the project archive to the Science Data Center(s). Details should be provided of the schedule, products and their volumes, and all the associated items.

3.3.2 Directories and Catalogs

This section should address metadata and the associated mechanisms for the identification and location of data sets and data analysis tools. The format of catalogs

and any browse products should be addressed in addition to the type of information that can be obtained.

3.3.3 Standards and Policies

Projects and investigators should archive data conforming to those standards and policies which will facilitate subsequent data access and use. This section of the DMP should describe which standards and policies for documentation, formats, and media will be used for the data to be archived on an overall basis. More detailed information for each data set should be deferred to an ICD and the ICD referenced here.

3.3.4 Networking Requirements

Networking requirements of the project for accomplishing data transfer should be summarized in this section, including bandwidth, latency and cybersecurity requirements.

4.0 Products

Products resulting from the project include standard science data products, and other associated archive products such as documentation and pre-launch calibration files. This section should describe what the project proposes to archive, and when and where it is going to be archived. This section provides more details on the items identified in section 4.3.

4.1 Science Data Product Summary

Science data products include data sets generated by the project. This section of the DMP should identify and describe the data sets expected to be generated. This includes the science data itself, associated ancillary data and aircraft navigation data. The basics to be covered include what data are going to be generated, when they are to be generated, and when they are to be archived.

The science data products for each of "n" instruments should be summarized in the following sections 4.1.1 through 4.1.n.

4.1.1 Instrument "A"

This section should summarize all science data products to be generated from instrument "A". Included with the products should be documentation for correct and independent use of the data. The expected range of use to be supported by this documentation should be indicated in the write-up, as well as any expected limitations. The following types of data sets should be included:

- Meta Data
- Low-level Processed Data
- High-level Processed Data
- Scientific Results
- Science algorithm source code
- Algorithm Theoretical Basic Documents (ATBD)
- Data quality documentation

A summary table should be provided for each instrument, including the planned data products.

4.2 Associated Archive Products

This section should include descriptions of archive products which are not included in the science data products of section 4.1. Discussion should include what is going to be saved, when it is going to be saved, where it will be saved, etc. If these products are associated with particular instruments or data sets, that information should be provided. The types of products which may be included in this section include the following:

- Operations histories

- Aircraft flight reports
- Analysis software
- Hardware descriptions
- Pre-launch test results
- Samples of test data used in algorithm testing
- Engineering drawings
- Navigation data
- Instrument geometry description
- Calibration data
- Validation data

All products which are expected to be archived should be identified.

5.0 Special Considerations

This section is available to address special considerations not covered in the standard structure provided in the above guidelines.

6.0 Acronyms

This section is for providing a glossary of acronyms used in the document.